

AD \_\_\_\_\_

GRANT NUMBER DAMD17-94-J-4383

TITLE: Developing and Implementing the AJCC Prognostic System  
for Breast Cancer

PRINCIPAL INVESTIGATOR: Philip H. Goodman, M.D.

CONTRACTING ORGANIZATION: University of Nevada  
Reno, Nevada 89520

REPORT DATE: August 1996

TYPE OF REPORT: Annual

19970314 067

PREPARED FOR: Commander  
U.S. Army Medical Research and Materiel Command  
Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;  
distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

# REPORT DOCUMENTATION PAGE

Form Approved

OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE August 1996	3. REPORT TYPE AND DATES COVERED Annual (1 Aug 95 - 31 Jul 96)	
4. TITLE AND SUBTITLE Developing and Implementing the AJCC Prognostic System for Breast Cancer			5. FUNDING NUMBERS DAMD17-94-J-4383	
6. AUTHOR(S) Philip H. Goodman, M.D.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Nevada Reno, Nevada 89520			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Commander U.S. Army Medical Research and Materiel Command Fort Detrick, Frederick, MD 21702-5012			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200)  The goal of this project is the creation of a computer-based prognostic system for breast cancer that: (1) is significantly more accurate than the TNM staging system, (2) that predicts survival over time based on therapy, and (3) presents its predictions in a manner that physicians can understand and use.				
14. SUBJECT TERMS American Joint Committee on Cancer, Artificial Neural Networks, Prognostic Factors, Staging, Outcomes, TNM, Breast Cancer, Survival, Recurrence			15. NUMBER OF PAGES 15	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

## FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

   Where copyrighted material is quoted, permission has been obtained to use such material.

   Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

   Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

(NA)    In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

   For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

(NA)    In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

(NA)    In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

(NA)    In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

 7/2/86  
PI - Signature Date

Introduction	3-4
Detailed Report by Task	5-8
Conclusions	9
References	10-11
Attachment # 1 - Accuracy methods	12
Attachment # 2 - Graphical output	13

## Introduction

The goal of this project is the creation of a computer-based prognostic system for breast cancer that: is significantly more accurate than the TNM staging system, predicts survival over time based on therapy, and presents its predictions in a manner that physicians can understand. This project can be viewed as consisting of three components: (1) Data analysis and prognostic factor evaluation, (2) developing the prognostic model, and (3) implementing a clinically useful system, i.e., breast cancer prognostic factors, the artificial neural network statistical model, and the clinician user interface.

The first year of research was characterized by work on the artificial neural network statistical model and related statistical models, specifically tasks 2.1.2 (artificial neural network generating survival curves), 2.1.3 (determining the accuracy of the survival curves), 1.03, 2.1.4 (comparing the accuracy of the artificial neural network to other statistical models), 2.2 (implementing an effective solution for missing data in training and performance), and 2.3 (dealing with censored data).

In addition, during the first year we started work related to data analysis and prognostic factors including 1.02 and 1.08.1 (recurrence as an endpoint), 1.04.2 (creating a taxonomy of prognostic factors in breast cancer), 1.04.3 (writing a book on prognostic factors in breast cancer, in preparation), 1.06.3 (determining minimum data set size), 1.11 (examining physician breast cancer survival estimates). We also began work on 3.1 (the code) and 3.2 (the physician interface).

Also during the first year we added three tasks, (1) a comparison of the two main American cancer data bases, namely, the Surveillance, Epidemiology, and End Results and the National Cancer Data Base data bases. (2) An examination of the issue of what to do when confronted with cases not lost completely at random and competing risks. (3) Computerization of the TNM staging system.

The second year of research was characterized by the continuation of work begun in the first year and by data analysis and prognostic factor development, specifically tasks 1.01 and 2.1.1 (extending the survival

endpoint from five to ten years), 1.02 and 1.03 (see first year), 1.05 (the identification of high risk node negative women), 1.06 (clinical trials), and 1.07 (therapy). Work continued on 2.1.2 (artificial neural network generating survival curves), 2.1.3 (creating a new method for assessing prediction accuracy), 1.03 and 2.1.4 (model comparisons). In addition, during the second year work began on 1.04.1 (new molecular-genetic prognostic factors). Work was completed on the computerization of the TNM staging system and the comparison of the two national cancer data bases.

## Detailed Report: by Task

## Task 1. Data analysis and prognostic factor evaluation.

## 1.01) Extend analysis of binary survival endpoint to 10 year survival.

We have analyzed SEER 10 year survival data. We found that the predictors collected at disease discovery are less accurate in predicting 10 year survival than 5 year survival.

**SEER 1977 - 1982 Breast Cancer Data: 10 year Survival Prediction Accuracy, TNM Variables**

PREDICTION MODEL	ACCURACY*	SPECIFICATIONS
TNM Stages	.692	Ø,I,IIA,IIB,IIIA,IIB,IV
Artificial Neural Network	.730 <sup>§</sup>	3-5-1

\* The receiver operating characteristic area.

<sup>§</sup>  $p < .01$

Five year survival accuracy for the TNM staging system was .720 and for the artificial neural network, .784.

## 1.02) Extend the analysis to recurrence as an endpoint.

Completed, in last year's report.

## 1.03) Comparison of prognostic models.

Completed, in last year's report.

## 1.04.1) New prognostic factors

We have obtained from Duke University a data set that contains, in addition to the TNM variables age, estrogen and progesterone receptor status, histology, p53, and erbB-2. The accuracy results are shown below. These new prognostic factors produce major increases in prognostic accuracy. This is a very encouraging result.

**Duke Breast Cancer Data: 5 year Survival Prediction Accuracy**

PREDICTION MODEL	ACCURACY*	SPECIFICATIONS
pTNM Stages	.567	Ø,I,IIA,IIB
Stepwise Logistic Regression	.865	no interactions
Backpropagation Neural Network	.926	9-5-1

\* The area under the curve of the receiver operating characteristic.

These results are based on slightly more than 100 cases, so the results must be interpreted cautiously. We will be obtaining several hundred additional cases from Duke in the near future. We will use the additional cases to confirm these results.

## 1.04.2) Create a taxonomy of prognostic factors in breast cancer.

Completed, in last year's report.

#### 1.04.3) Prognostic Factors in Breast Cancer book.

The book is being prepared for publication in 1997. *Burke HB, Henson DE. Prognostic Factors and Systems in Oncology. Kluwer Academic Publishers Inc., in preparation.*

#### 1.05) High risk node negative women

Based on our initial results with the new prognostic factors contained in the Duke data set (1.04.1) we believe that some of these factors will be very useful in identifying high risk node negative women. This work is ongoing.

#### 1.06.1, 1.06.2) Clinical trials

The increases in prognostic accuracy we described last year and this year suggest that we are finding more homogeneous patient populations. The Duke data set contains treatment information and we are currently using it to discover which new prognostic factors predict response to specific therapies. This work is ongoing.

#### 1.06.3) Determining minimum data set size.

Completed, in last year's report.

#### 1.08.1) Recurrence analysis.

Completed, in last year's report.

#### 1.11) Patient information and physician credibility.

Completed, in last year's report.

#### Task 2. Developing the prognostic model.

##### 2.1.1) Generate survival curves for 10 year data.

Refer to 1.01 for details.

##### 2.1.2) Generating survival curves.

Completed, in last year's report.

##### 2.1.3) Determining the accuracy of the survival curves.

We are currently working on a new measure of prediction accuracy that we call "A". It includes the area under the receiver operating characteristic as a special case. This work is ongoing. We have implemented several accuracy methods and derived the asymptotic variance for each in order to assess the adequacy of each method. (See Attachment).

##### 2.1.4) Comparison of artificial neural networks with Cox proportional hazards model.

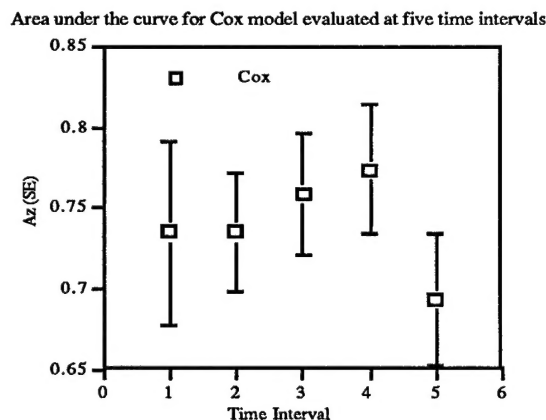
We began our comparison of the Cox by examining whether breast cancer violates the proportional hazards assumption of the model. Proportional hazards methods include the Cox (1972), and less commonly the Weibull or exponential distributions (Evans, 1993). Proportional hazards methods assume that the hazard



of each patient is proportional to the hazards of all the other patients and that a individual patient's hazard is related to that patient's relative risk. The Cox model does not create survival curves. For Cox-related survival curves a baseline hazard must be introduced (Breslow-Cox estimates; Breslow, 1974). Some researchers incorrectly believe that only regression methods that assume proportional hazards can deal with censoring, but a multiinterval regression model that drops patients during the interval in which they are censored is capable of dealing with censoring. It is always vital to test the proportional hazards assumption when using a regression method that relies on it. There are several methods for assessing proportional hazards violation, including Schoenfeld's partial residuals (Schoenfeld, 1982) and the log hazard ratio as a function of time (Gore, 1986). We have created a method somewhat similar to Gore. We construct a Cox model, divide the time into sub-intervals, and assess the accuracy of the model for each sub-interval. If proportional hazards holds, accuracy should be constant across sub-intervals. Results for breast cancer are shown below.

**TABLE.** Area under the receiver operating characteristic (Az) for two Cox models; breast cancer (five one-year intervals) N = 1,222 and melanoma (three six-month intervals) N = 60.

Model/Interval	1	2	3	4	5
Breast (SE)	.734 (.057)	.735 (.036)	.758 (.038)	.773 (.040)	.693 (.041)



Because the Az values are not constant across the sub-intervals, proportional hazards does not hold for breast cancer.

Artificial neural networks are a general regression method that do not assume proportional hazards and can capture nonlinearity and complex interactions (Burke, 1994, 1995a). Multiinterval artificial neural networks can handle censoring in the same way that multiinterval logistic regression models handle censoring. It seems clear that proportional hazards is probably not appropriate for breast cancer or lung cancer (results not reported).

## 2.2) Missing data (see discussion in last year's report also)

We have completed work on an efficient missing data mechanism. It is available for use by interested researchers.

*Rosen DB. Burke HB. Applying a gaussian-bernoulli mixture model network to binary and continuous missing data in medicine. Submitted for publication.*

*Burke HB (ed). Missing Data: Advanced Methods and Models. M.I.T. Press, in discussion.*

### 2.3) Censored cases

Discussed in detail in last year's report (2.1.2).

### Task 3. Implementation of a clinically useful prognostic system

#### 3.1) Computer code for artificial neural networks.

All our work is written in either C, C++, or XLISP-STAT.

#### 3.2) Physician interface.

It is very important that physicians find the new prognostic system easy to use and useful. To this end we have implemented the prognostic system on a DOS platform with a Windows interface. We are presenting the system to clinicians and receiving feedback regarding what is important to them in terms of information and the graphical display of the information. (see 2.1.2) See Attachment for screen output.

#### Tasks added to the project

(1) The computerization of the TNM staging system for breast cancer has been completed and has been integrated into the prognostic system.

(2) Comparison of the NCDB and the SEER data sets in terms of breast cancer.

We have completed our comparison of the two national breast cancer data bases, the National Cancer Data Base (NCDB) and its associated Patient Care Evaluation (PCE), and the Surveillance, Epidemiology, and End Results (SEER) data sets. We evaluated them in terms of: (i) representativeness, is the data set an unbiased representation of the breast cancer population. (ii) Incidence/prevalence, how good is the data set in capturing the incidence and prevalence of breast cancer. (iii) Prognosis/outcome, how good is the data set in providing information that is useful for predicting outcome. An overview of the results are shown below.

	SEER	NCDB
Representativeness	good	good
Incidence/prevalence	good	adequate
Prognosis/outcome	not acceptable	adequate

Both are representative of the breast cancer population. SEER does a better job at incidence and prevalence because it ascertains all cases in a catchment area, regardless of whether the hospital belongs to the American College of Surgeons accreditation program, and it contains relatively little missing data. NCDB contains a great deal of missing data. SEER can not be used for prognosis because it does not provide therapy data due to the unreliability of their data. NCDB suffers from a lack of follow-up, resulting in high censoring rates.

*Burke HB, Hoang A, Visintainer, P. Comparison of the two national cancer data sets: SEER and NCDB. In preparation.*

## Conclusions

The second year of research was characterized by the continuation of work on the artificial neural network statistical model and related statistical models and by data analysis and prognostic factor development. In summary, the research is going well we are ahead of our time schedule. We feel that we will be able to successfully meet our goal of providing a computer-based prognostic system that is more accurate than the TNM staging system and that is easy to use and understand within the four year time frame of this grant. In addition, we have created several new systems that we believe will advance the domain of cancer prognosis, e.g., artificial neural network survival-over-multi-interval-time models, an effective missing data method for training and performance, and a new approach to the assessment of prediction accuracy.

## References

### Publications and Presentations Related to this Grant During The Second Funding Year

#### **Books**

- Burke HB, Henson DE. Prognostic Factors and Systems in Oncology. Kluwer Academic Publishers, in preparation.
- Burke HB (ed). Missing Data: Advanced Methods and Models. M.I.T. Press, in discussion.
- Burke HB (ed). Artificial Neural Networks in Medicine. Kluwer Academic Publishers, in preparation.

#### **Peer Reviewed Book Chapters**

- Burke HB, Rosen DB, Goodman PH. Comparing the prediction accuracy of artificial neural networks and other statistical models for breast cancer survival. In G. Tesauro, D.S. Touretzky, T.K. Leen, eds. Advances in Neural Information Processing Systems 7. Cambridge, MA; MIT Press, 1995, 1063-67.
- Burke HB. Statistical analysis of complex systems in biomedicine. In D. Fisher and H. Lenz, eds. Learning from Data: Artificial Intelligence and Statistics V. New York: Springer-Verlag, 1996, 251-258.
- Burke HB. Integrating multiple clinical tests to increase predictive accuracy. In M. Hanausek, Z. Walaszek, eds. Methods in Molecular Biology: Tumor Marker Protocols. Humana Press Inc., 1996, in press.

#### **Peer Reviewed Journals and Proceedings**

- Burke HB, Rosen DB, Goodman PH. Comparing the prediction accuracy of statistical models and artificial neural networks in breast cancer. Preliminary Papers of the Fifth International Workshop on Artificial Intelligence and Statistics 1995; 87.
- Burke HB, Goodman PH, Rosen DB. Applying artificial neural networks to medical knowledge domain. Proceedings of the International Symposium on Integrating Knowledge and Neural Heuristics 1995, in press.
- Burke HB, Goodman PH, Rosen DB. A computerized prediction system for cancer patient survival that uses an artificial neural network. Proceedings of the First World Congress on Computational Medicine and Public Health 1995, in press.
- Burke HB. Artificial neural networks and biomedicine. Proceedings of the Workshop on Environmental and Energy Applications of Neural Networks 1995, in press.
- Burke HB. The importance of artificial neural networks in biomedicine. Proceedings of the World Congress on Neural Networks. Hillsdale, NJ: Lawrence Erlbaum Associates Inc. 1995, 725-30.
- Burke HB, Hoang A, Rosen DB. Survival function estimates in cancer using artificial neural networks. Proceedings of the World Congress on Neural Networks. Hillsdale, NJ: Lawrence Erlbaum Assoc. Inc. 1995, 742-7.
- Burke HB, Goodman PH, Rosen DB, Henson DE, Weinstein JN, Hellier JH, Harrell Jr. FE, Marks JR, Winchester DP, Bostwick DG. Artificial neural networks improve cancer survival prediction accuracy. Cancer, in press.
- Rosen DB, Burke HB, Goodman PH. Improving prediction accuracy using a calibration postprocessor. Proceedings of the World Congress on Neural Networks. Hillsdale, NJ: Lawrence Erlbaum Assoc. Inc. 1996, in press.
- Burke HB. Measuring classification/prediction accuracy. Proceedings of the World Congress on Neural Networks. Hillsdale, NJ: Lawrence Erlbaum Assoc. Inc. 1996, in press.

- Rosen DB, Burke HB. Applying a gaussian-bernoulli mixture model network to binary and continuous missing data in medicine. Submitted for publication.
- Burke HB. The TNM staging system. In preparation.
- Hoang A, Burke HB. M: A measure of discriminative accuracy. In preparation.
- Burke HB, Henson DE, Bostwick DG, Hoang A. Backward and forward relationships between prior events and disease. In preparation.
- Hoang A, Burke HB. Methods for dealing with cases nonrandomly lost-to-follow-up. In preparation.
- Burke HB, Hoang A. Assessment of multi-interval event models. In preparation.
- Burke HB, Hoang A, Visintainer P. Comparison of the two national cancer data sets: SEER and NCDB. In preparation.

### Invited papers

- Burke HB. Defining the computer-aided diagnostic device domain. Proceedings of the World Congress on Neural Networks. Hillsdale, NJ: Lawrence Erlbaum Assoc. Inc. 1996, in press.

### Presented papers

- Burke HB. Outcome prediction in cancer. 1995 International Conference on Health Policy Research: Methodologic Issues in Health Services and Outcome Research, Harvard Medical School, Boston MA, December 2 - 3, 1995.
- Burke HB. Defining the Computer-aided Diagnostic Device Domain. U.S. Food and Drug Administration Computer-aided Diagnostic Device Workshop, Rockville MD, January 26, 1996.
- Burke HB. Computer-based Clinical Decision Support System in Oncology. The Eli Lilly Lecture, Conference on Prognostic Factors and Rational Treatment of Cancer, Yorkshire Cancer Organization, Leeds UK, July 3, 1996.
- Burke HB. Issues in the regulation of medical software devices. Plenary session, FDA/NLM Software Policy Workshop, National Institutes of Health, Bethesda MD, September 3 - 4, 1996.
- Rosen DB, Burke HB, Goodman PH. Improving prediction accuracy using a calibration postprocessor. World Congress on Neural Networks, San Diego CA, September 15 - 20, 1996.
- Burke HB. Measuring classification/prediction accuracy. World Congress on Neural Networks, San Diego CA, September 15 - 20, 1996.
- Burke HB. Defining the computer-aided diagnostic device domain. World Congress on Neural Networks, San Diego CA, September 15 - 20, 1996.

### Conference positions

- 1995 *Scientific Advisor*. Cancer Prognostic Factors Conference, Cambridge Healthtech Institute, Arlington, VA, June 7 - 8, 1995.
- 1995 *Program Committee and Co-chair*. Biomedical Applications of Neural Networks section, World Congress on Neural Networks and 1995 International Neural Network Society Annual Meeting, Washington DC, July 17 - 21, 1995.
- 1995 *Workshop Chair*, Missing Data: Methods and Models. Neural Information Processing Systems 1995, Vail CO, December 1, 1995.
- 1996 *Program Committee and Co-chair*, Biomedical Applications Section, World Congress on Neural Networks and 1996 International Neural Network Society Annual Meeting, San Diego CA, September 15 - 20, 1996.
- 1996 *Workshop Chair*, Model Accuracy: Issues and Methods. Neural Information Processing Systems 1996, Vail CO, December 2, 1996.
- 1996 *Co-organizer*, p53: A Prognostic Factor in Cancer. National Cancer Institute, in preparation for Spring, 1997, Bethesda, MD.

MEASURE OF ASSOCIATION	PROBABILITY STATEMENT	INTERPRETATION	ESTIMATION (DISCRETE)	ASYMPTOTIC VARIANCE OF THE ESTIMATE (DISCRETE)
Goodman and Kruskal's Gamma	$\frac{P[(x_1 > x_2, y_1 > y_2) \text{ or } (x_1 < x_2, y_1 < y_2)] - P[(x_1 > x_2, y_1 < y_2) \text{ or } (x_1 < x_2, y_1 > y_2)]}{P[(x_1 > x_2, y_1 > y_2) \text{ or } (x_1 < x_2, y_1 < y_2)] + P[(x_1 > x_2, y_1 < y_2) \text{ or } (x_1 < x_2, y_1 > y_2)]}$	The excess of the degree of the same movement in X and Y over the degree of the opposite movement in X and Y, after removing ties in $X_1$ and $X_2$ and $Y_1$ and $Y_2$ .	$\frac{P-Q}{P+Q}$	$\frac{16 \sum_{x,y} N_{xy}(QA_{xy} - PD_{xy})^2}{(P+Q)^4}$
Somer's $D_{x:y}$	$\frac{P[(x_1 > x_2, y_1 > y_2) \text{ or } (x_1 < x_2, y_1 < y_2)] - P[(x_1 > x_2, y_1 < y_2) \text{ or } (x_1 < x_2, y_1 > y_2)]}{P(y_1 \neq y_2)}$	The excess of the degree of the same movement in X and Y over the degree of the opposite movement in X and Y, after removing ties in $Y_1$ and $Y_2$ .	$\frac{P-Q}{N^2 - \sum_y N_y^2}$	$\frac{4}{W_y^2} \sum_{x,y} N_{xy} [w_x d_{xy} - (P-Q)(N-N_y)]^2$
Kendall's Tau-b	$\frac{P[(x_1 > x_2, y_1 > y_2) \text{ or } (x_1 < x_2, y_1 < y_2)] - P[(x_1 > x_2, y_1 < y_2) \text{ or } (x_1 < x_2, y_1 > y_2)]}{\sqrt{P(x_1 \neq x_2)P(y_1 \neq y_2)}}$	Tau-a is the excess of the degree of the same movement in X and Y over the degree of opposite movement in X and Y. Tau-b is Tau-a relative to the geometric average of the chance of no ties in $X_1$ and $X_2$ and the chance of no ties in $Y_1$ and $Y_2$ .	$\frac{P-Q}{\sqrt{\left(N^2 - \sum_x N_x^2\right)\left(N^2 - \sum_y N_y^2\right)}}$	$\frac{1}{W^2} \left[ \sum_{x,y} N_{xy} (2w_x d_{xy} + [\text{tau}_b] v_{xy})^2 - N^3 [\text{tau}_b]^2 (w_x + w_y)^2 \right]$
Area under the receiver operating characteristic ( $A_2$ )	$P(x_1 > x_2   y_1 > y_2), \text{ continuous } x$ $P(x_1 > x_2   y_1 > y_2) + \frac{1}{2} P(x_1 = x_2   y_1 > y_2), \text{ discrete } x$	Given the direction of movement of Y, the degree to which X has the same direction of movement.	$\frac{5(P+Z)}{L_y}$	$\frac{1}{N} \left[ \sum_{x,y} N_{xy} \theta_{xy}^2 - \left( \sum_{x,y} \frac{N_{xy}}{N} \theta_{xy} \right)^2 \right]$

$N_{xy}$  = number of pairs (X, Y) such that  $X=x$  and  $Y=y$ ,  $N = \sum_{x,y} N_{xy}$ ,  $N_{x \cdot} = \sum_y N_{xy}$ ,  $N_{\cdot y} = \sum_x N_{xy}$ ,  $P = \sum_{x,y} N_{xy} y_1 \sum_{x_1 > x_2} N_{x_1 y_2}$ ,  $Q = 2 \sum_{x_1 > x_2} N_{x_1 y_1} \sum_{x_1 < x_2} N_{x_2 y_2}$ ,

$$L_x = \sum_{x_1} N_{x_1} \cdot \sum_{x_2} N_{x_2} \cdot, L_y = \sum_{y_1} N_{y_1} \cdot \sum_{y_2} N_{y_2} \cdot, Z = \sum_{x,y} N_{xy} y_1 \sum_{x_1 > x_2} N_{x_1 y_2} A_{xy} = \sum_{x_1 > x_2} N_{x_1} y' + \sum_{x_1 < x_2} N_{x_1} y', D_{xy} = \sum_{x_1 > x_2} N_{x_1} y' + \sum_{x_1 < x_2} N_{x_1} y', d_{xy} = A_{xy} - D_{xy}, W_x = N^2 - \sum_x N_x^2,$$

$$W_y = N^2 - \sum_y N_y^2, W = \sqrt{W_x W_y}, v_{xy} = N_{x \cdot} \cdot W_y + N_{\cdot y} \cdot W_x, \theta = \frac{1}{2L_y^2} [L_y (2A_{xy} + N_{x \cdot} - N_{xy}) - (P+Z)(N-N_y)],$$

$$\eta_{xy} = \frac{1}{[L_y(N^2 - L_y)]^2} \left\{ N^3 \left[ A_{xy} L_y - \frac{P}{2} (N - N_{\cdot y}) \right] + N^2 \left[ P(N - N_{\cdot y}) - (N - N_{x \cdot}) L_y^2 - A_{xy} L_y^2 \right] + N \left[ (N - N_{x \cdot}) L_y^3 - L_x (N - N_{\cdot y}) L_y^2 \right] \right\}$$

# Patient Data

Patient Name	<input type="text"/>	Tumor	<input type="text"/>
Patient ID	<input type="text"/>	Tumor Size	<input type="text" value="2"/>
Institution	<input type="text"/>	Lymph Nodes Positive	<input type="text" value="3"/>
Date	<input type="text"/>	Lymph Nodes Examined	<input type="text" value="10"/>
Physician	<input type="text"/>	Distant Metastasis	<input type="text" value="n"/>
Cancer Type	<input type="text" value="Breast"/>	Estrogen Receptor	<input type="text" value="n"/>
Prediction Type	<input type="text" value="5 Year Survival Curve"/>	Progesterone Receptor	<input type="text" value="n"/>
		Menopausal Status	<input type="text" value="y"/>
		Grade	<input type="text" value="1"/>
		Age	<input type="text" value="55"/>
		Lymph Nodes pTNM	<input type="text" value="1"/>

Predict

Multi Predict

Clear

Save

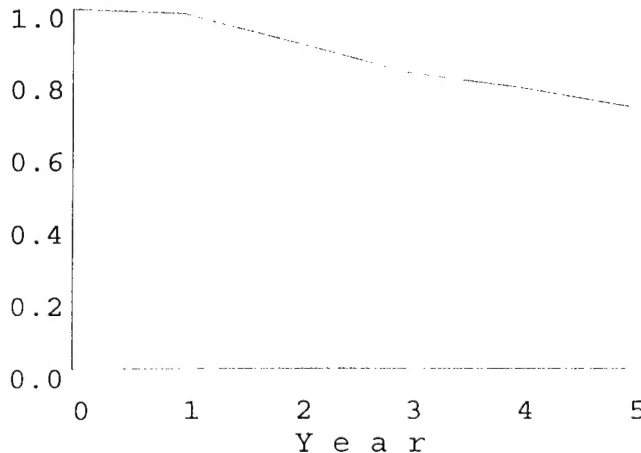
<output screen>

Patient Name:  
Patient ID:  
Institution:  
Date:  
Physician:  
Cancer Type:  
Prediction Type:

Factors: tsize =  
Tumor:  
T Size: 2.4 LN Pos: 4  
LN Exam: 10 Mets: n  
ER: n PR: n  
Menopausal: n Grade: 2  
Age: 40 LN pTNM: 2

Year	ANN Prediction
0	1.000
1	0.983
2	0.898
3	0.82
4	0.775
5	0.721

P  
r  
o  
b  
a  
b  
i  
l  
i  
t  
y



TNM Stage: IIIA

Prediction: 0.595